BAYESIAN MODEL SELECTION OF STRUCTURAL EXPLANATORY MODELS

Bahar Dadashova* INSIA-UPM, Spain
Blanca Arenas, INSIA-UPM, Spain
Jose Mira McWilliams, INSIA-UPM, Spain
Francisco Aparicio, INSIA-UPM, Spain
* bahar@etsii.upm.es

Introduction

In this study we are proposing a Bayesian model selection methodology, where the best model from the list of candidate structural explanatory models is selected.

The model structure is based on the Zellner's (1971) explanatory model with autoregressive errors.

For the selection technique we are using a parsimonious model, where the model variables are transformed using Box and Cox (1964) class of transformations.

Structural model

The following structural explanatory model with AR(2) error term is used:

$$y_t^{(\lambda_y)} = \sum_{k=1}^{K} \beta_k x_{kt}^{(\lambda_x)} + u_t$$

$$u_t = \sum_{l=1}^{r} \rho_l u_{t-1} + w_t$$

β_k are the regression coefficients, u_t is an error term with the AR(2) structure and w_t are assumed to be white noise, N(0, σ_w^2).

$$y_t^{(\lambda_y)} = \begin{cases} \frac{y_t^{\lambda_y}-1}{\lambda}, if\ \lambda \neq 0 \\ ln(y_t), if\ \lambda = 0 \end{cases}$$

Variables are Box-Cox transformed, with BCT coefficient λ_Y.

Bayesian model selection

Methodology :

- Stage 1- Selection of explanatory variables: A set of two-input models (TIM), using 28 explanatory variables (378 models) were built and estimated. The explanatory variables belonging to the same TIMs are transformed with the same value of λ_X. There are three fixed values of λ_X= (0.5, 0.1, 0.5). Explanatory variables that appear in the first 50 models (for each value of λ_X) with the best goodness of measures are selected.

- Stage 2- Estimation of BCT for the response: Using the explanatory variables selected in Stage 1 optimize the BCT value for the dependent variable with respect to λ_Y.

- Stage 3- Estimation of BCT for the explanatory variables: Build a model using the explanatory variables selected in the first stage, where both the dependent and independent variables are power transformed. BCT for dependent variables is obtained at Stage 2. The variables belonging to the same group (G) were transformed with the same value of λ_X, where λ_X= (0.5, 0.1, 0.5). A total of 243 models are estimated.

- Stage 4- Final model selection: The final model selection is based on expected signs of the explanatory variables selected in Stage 1 and goodness of fit measures (DIC).

- Stage 5- Cross-validation through prediction: Model selection procedure is cross-validated by prediction analysis.

Data:

The analysis was carried out using the monthly data on van involved traffic accidents in Spain. The data covers the period of 2000-2011. Road safety measure is the number of van involved fatal accidents.

Results

	VARIABLES	ESTIMATE	ELASTICITY	λ_X
G1	Vehicle km travelled	6.353	(+) 11.1%	0.1
	Fuel consumption	6.584	(+) 11.8%	0.1
G2	Vehicle age>10 y.o.	2.256	(+) 0.3%	-0.5
	Total employment	0.006	(+) 0.6%	0.5
G3	Maintenance	-0.059	(-) 1.4%	0.5
	Fuel price	-0.332	(-) 0.7%	0.5
	Alcohol controls	-0.000014	(-) 0.01%	0.5
G4	Radar checks	-0.000048	(-) 0.08%	0.5
	Driving license suspended	-0.018	(-) 1.8%	0.5
G5	Length of highways	-43.740	(-) 9.6%	0.5
G6	Penalty Point System	-0.244	(-) 1.1%	NA
	PARAMETER ESTIMATES			
	ρ_1	0.219		
	ρ_2	0.154		
	τ_w	9.437		
	GOODNESS OF FIT STATISTICS			
Deviance Information Criteria		93.747		
Pseudo-R²		93%		

Table 1. The explanatory variables are divided into groups (G), where each group is assigned a different BCT coefficient, λ_X. Elasticity is a percentage increase (+) /decrease (-) in response as a result of 10% increase in regressors; ρ_1 and ρ_2 are autoregressive error term; τ_w is error variance.

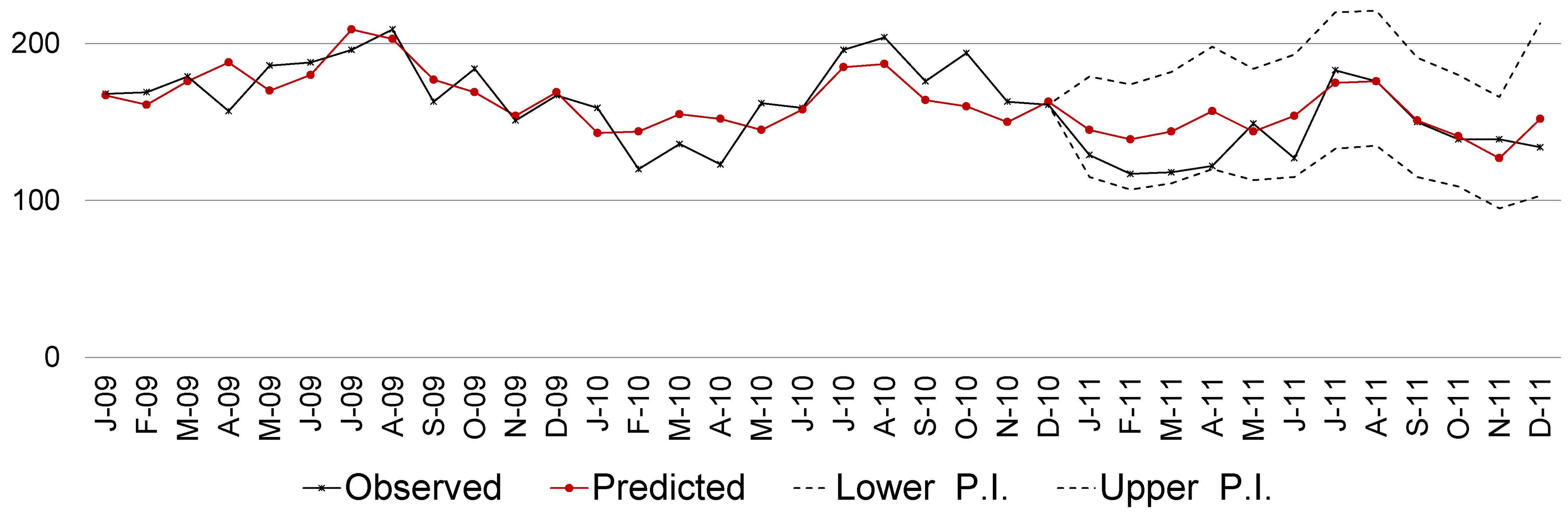

Figure 1. Bayesian predictions for van involved fatal accidents: ACCMOR, 2009-2011.

Conclusions

- Unlike the usual structural explanatory models applied in road safety the model presented is less parsimonious.
- The proposed strategy allows the consecutive estimation of several models at once thus making the model estimation and selection process more efficient and less time consuming compared to DRAG models.
- The results of the Bayesian estimation conform to the results obtained in previous empirical studies on road safety analysis.
- The predictions show that the estimated prediction interval is valid for all of the observations and closely follows the observed time series.

Main references

1. Gaudry, M. and Lassarre, S. (2000). Structural road accident models. The international DRAG family. Oxford: Elsevier Science.
2. Raftery, A. E. (1995). Bayesian model selection in social research. Sociological Methodology, 25, 111-164.
3. Robert, C. P., and Casella, G. (2004). Monte Carlo statistical methods (Vol. 319). New York: Springer.
4. Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics. J. Wiley and Sons, New York.